KW

library

METEOROLOGICAL OFFICE
139416
31 JAN 1983
LIBRARY

MET O 11 TECHNICAL NOTE NO. 164

A FULLY IMPLICIT LAGRANGIAN METHOD AS AN ALTERNATIVE APPROACH TO NUMERICAL WEATHER PREDICTION

by

R J Purser*

Meteorological Office (Met O 11)
London Road
Bracknell
Berkshire
U.K.

January 1983

N.B. This paper has not been published. Permission to quote from it must be obtained from the Assistant Director of the above Meteorological Office Branch.

* On secondment at the Space Science and Engineering Center, University of Wisconsin, Madison, Wisconsin 53706 (from 1st February 1983).

FH2A

MET O 11 TECHNICAL NOTE NO 164

A FULLY IMPLICIT LAGRANGIAN METHOD AS AN ALTERNATIVE APPROACH TO NUMERICAL WEATHER PREDICTION.

1. Introduction

The conventional formulation for a grid point model involves holding the horizontal coordinates of the grid fixed in space and representing the equations of the fluid flowing past the grid in Eulerian form. In spite of various attempts (eg Crowley (1968), Gadd (1978)) to aleviate the disruption of structure caused by the numerical treatment of advection, the rapid destruction of detail on the smallest resolvable scales remains an almost unavoidable consequence of the numerical representation of advection on a fixed grid, as is apparent from Figure 1. The dynamical consequences of this erroneous disruption are largely unknown but could conceivably be severe on all scales resolved by a few days for a synoptic scale model. An alternative approach, largely neglected in recent years, is to use a grid which is itself advecting with the flow and to compute the trajectories of each grid point from the forces implied by the disposition of its neighbours.

This has been undertaken with apparent success for an amorphous ensemble of points by Mesinger (1971), whilst semi-Lagrangian methods (which still rely to a large extent on a fixed grid) have been attempted for simplified formulations of the dynamical equations, for example, Wiin-Nielsen (1959) and Sawyer (1963). Their results compared favourably with Eulerian methods. Recently, Robert (1981), (1982) has attempted a semi-Lagrangian algorithm to simulate a 'shallow water' system on the sphere using semi-implicit methods to retain numerical stability with a time step of two hours (considerably greater than that permitted in the more conventional methods). His results appear very encouraging with numerical errors attributable to the relatively long timesteps still remaining small. However, apart from this work there appears to be little active research into the use of Lagrangian integration methods in meteorology although in other branches of fluid simulations Lagrangian methods have proved their value, often in the form of vorticity codes (eg Christiansen (1970), (1973)). While research into Lagrangian methods has declined since the emergence of primitive equation models

the performance of the best of these Eulerian models remains somewhat disappointing considering their resolution and complexity. It seems that the resolution to which structures may be faithfully dealt with in the fixed grid models is far short of the resolution that the grid is theoretically capable of. It is important to realise that this applies also in the vertical where a poor representation of advection must presumably lead to a very serious misrepresentation of geostrophic adjustment - essentially the process of differential vertical motion (implying vortex stretching) that sustains the nice balance observed between the wind and pressure fields and governs the degree of vertical coherence of features as well as their rate of development.

It therefore seems timely to reconsider the possibility of applying Lagrangian methods in a way carefully formulated to overcome the intrinsic shortcomings of the Eulerian technique without incurring any more computation that is absolutely necessary to achieve numerical stability and the desired accuracy. Our concern will be mainly with the treatment of dynamics since this is where most problems are likely to arise. However, it must be emphasised that one of the chief motivations for considering a Lagrangian approach is that it seems to be the most natural way of simultaneously handling any large set of physical fields whose evolution is dominated by advection. This is certainly the case in an attempt to model the detailed development of weather, either for the purpose of operational forecasting or for simulating the processes whose statistical behaviour constitutes climate. Except for certain small scale features forced by topographic contrasts, the Lagrangian framework is especially effective at slowing the rates of change of the principal dynamical and bulk-physical variables, suggesting the important numerical advantage that for the majority of processes one might get by with considerably longer timesteps. In particular, through the use of implicit techniques the modeller may be finally freed from the unreasonably short timesteps imposed by the advective CFL criterion.

Obvious disadvantages of the Lagrangian formulation spring to mind: the evaluation of spatial derivations commonly used in the governing equations is

not such a direct process and constitutes an additional computional penalty for each such evaluation. A more serious objection is that a truly Lagrangian grid will rapidly distort beyond the limit of its usefulness as a coordinate system suitable for the accurate numerical representation of the physical processes. This problem is clearly most acute at the smallest resolvable scales of motion and requires a periodic renewal or 'rezoning' of the grid in order to avoid a numerical disaster. Rezoning essentially entails the defining of a new regular grid and the accurate interpolation of the data stored on the old grid to the new grid points. Many standard Eulerian advection schemes may be thought of as quasi-Lagrangian methods involving a simple rezoning to the original grid at every timestep. In a Lagrangian model it is not usually necessary to rezone every timestep but it must be done every few steps. Since the Lagrangian time-step will usually be several times larger than the equivalent Eulerian limit set by the CFL criterion we may expect that a Lagrangian grid rezoning operation occurs only for a considerable number of time steps of a comparable Eulerian model, and for this duration the explicit calculations of advective terms are entirely avoided. We should therefore be able to justify lavishing a great deal of care on the interpolation process to ensure that the consequent degradation of information is minimised.

A further problem with the straight-forward Lagrangian approach concerns the accuracy of the large scale structures. In a well constructed Eulerian coordinate system there usually exists a high degree of regularity and smoothness which is not guaranteed in the Lagrangian grid once the small scale motions have begun to deform it. It is due to the intrinsic smoothness of the Eulerian grids that large scale fields can be very accurately differentiated (implying that, except through non-linear interactions, large scale motions should be simulated faithfully). The small scale twisting and buckling of the Lagrangian grid will very probably jeopardise the accurate representation of the larger scales even when rezoning occurs well before the onset of any actual numerical

catastophe. It is never expected that a numerical simulation is particularly accurate near the limit of grid resolution but it would be ironical if the presence of these small scale motions were to spoil the simulation of the other scales of flow also.

These problems have been given some consideration and the proposals outlined in the following sections are intended to provide a reasonably logical guide for the construction of Lagrangian models structured to overcome the intrinsic difficulties just mentioned. While it is not intended to convey the impression that the application of Lagrangian methods will lead to simpler coding it is expected that the Lagrangian method should exhibit high standards with respect to the principal desiderata of numerical simulation, viz:

(i) Stability

(ii) Accuracy

(iii) Computational economy

The extra complexity compared with typical Eulerian code should, in the long run, be adequately compensated for by the ability of the proposed Lagrangian method to be easily generalised to arbitrary order of accuracy in space and time differencing and by the facility with which quasi-passive quantities may be advected.

2. Dynamics in a Lagrangian framework

We shall assume that a point P in space at latitude ϕ north, longitude θ east and height x_3 is described by the Cartesian coordinates of its projection in a polar stereographic frame,

$$P \equiv (x_1, x_2, x_3)$$

where
$$x_1 = \tan\left(\frac{\frac{\pi}{2} - \phi}{2}\right)\cos\theta$$

$$x_2 = \tan\left(\frac{\frac{\pi}{2} - \phi}{2}\right)\sin\theta \qquad (1)$$

The equator is mapped to the unit circle, $x_1^2 + x_2^2 = 1$, the Greenwich meridian to the line, $x_2 = 0$, $x_1 \geq 0$ as shown in Figure 2. The use of an orthomorphic (or conformal) projection has obvious advantages when dealing with vectors. Small distances ds are given by the metric formula:

$$ds^2 = \frac{1}{a^2}\left(dx_1^2 + dx_2^2\right) + dx_3^2 \qquad (2)$$

where a is the map factor,

$$a = \frac{1}{r_e}\frac{1}{(1+\sin\phi)} = \frac{1}{2r_e}\left(1 + x_1^2 + x_2^2\right) \qquad (3)$$

r_e is the radius of the earth.

An equally valid representation of P more suitable for locations in the southern hemisphere is:

$$x_1' = \tan\left(\frac{\frac{\pi}{2}+\phi}{2}\right)\cos\theta$$

$$x_2' = -\tan\left(\frac{\frac{\pi}{2}+\phi}{2}\right)\sin\theta \qquad (4)$$

The transformation between these choices is simply:

$$\omega = \frac{1}{\omega'}$$

where $\omega = x_1 + ix_2$ (5)

$\omega' = x_1' + ix_2'$

We quote without proof some geometrical results valid for a surface viewed in orthmorphic coordinates. The 2-dimensional gradient of a scalar, S is given in x_1, x_2 coordinates:

$$\left(\underset{\sim}{\nabla} S\right)_i = a\frac{\partial S}{\partial x_i} \qquad (6)$$

Under parallel transport by an infinitesimal displacement, dx_i the vector V_i changes its components according to:

$$dV_i = \sum_{j,k,l=1}^{2} \frac{dx_j}{a} \varepsilon_{jk} \Lambda_k \varepsilon_{il} V_l \qquad (7)$$

in which the factors representing the effect of the curvature of the coordinates are the vector components:

$$\Lambda_k = \left(\underset{\sim}{\nabla} (\log a) \right)_k = \frac{\partial a}{\partial x_k} \qquad (8)$$

and $\varepsilon_{12} = -\varepsilon_{21} = 1 \; ; \quad \varepsilon_{11} = \varepsilon_{22} = 0$

It follows that the covariant derivative of a vector field, $\underset{\sim}{V}$ is:

$$\left(\underset{\sim}{\nabla} \underset{\sim}{V} \right)_{ij} = a \frac{\partial V_j}{\partial x_i} - \sum_{k,l=1}^{2} \varepsilon_{ik} \Lambda_k \varepsilon_{jl} V_l \qquad (9)$$

In the case of polar stereographic coordinates these expressions are particularly simple, as, from (3) and (8):

$$\Lambda_k = \frac{x_k}{r_e} \qquad (10)$$

Using (7) and Newton's third law, a particle influenced by a net horizontal force-per-mass, $\underset{\sim}{F}$, changes its velocity components at the rate,

$$\frac{dU_i}{dt} = F_i + \sum_{j,k,l=1}^{2} U_j \varepsilon_{jk} \Lambda_k \varepsilon_{il} U_l \qquad (11)$$

where the velocity,

$$U_i = \frac{1}{a} \frac{dx_i}{dt} \qquad (12)$$

From

$$\frac{da}{dt} = \sum_{j=1}^{2} \frac{\partial a}{\partial x_j} \frac{dx_j}{dt}$$

we now obtain

$$\left.\begin{aligned} \frac{d^2 x_i}{dt^2} &= a\left(F_i + \sum_j \Lambda_j U_j U_i + \sum_{jkl} U_j \varepsilon_{jk} \Lambda_k \varepsilon_{il} U_l\right) \\ &\equiv a\left(F_i + G_i\right) \end{aligned}\right\} \quad (13)$$

where $\underset{\sim}{G}$ reduces explicitly to

$$\left.\begin{aligned} G_1 &= \Lambda_1\left(U_1^2 - U_2^2\right) + 2\Lambda_2 U_1 U_2 \\ G_2 &= 2\Lambda_1 U_1 U_2 - \Lambda_2\left(U_1^2 - U_2^2\right) \end{aligned}\right\} \quad (14)$$

In a Lagrangian description of atmospheric motion the terms $\underset{\sim}{F}$ correspond to the resultant of the pressure gradient, Coriolis and frictional forces and, as a cursory scale analysis will verify, these contributions are generally much larger than the 'geometric' terms, $\underset{\sim}{G}$.

We introduce Lagrangian coordinates, X_1, X_2, X_3 and assume that each is a sufficiently smooth function of x_1, x_2, x_3 to allow differentiation as often as is required. By definitions, these coordinates do not change following a particle. It is natural to choose one of these coordinates, say X_3 to increase from 0 at the terrain surface to some uniform maximum value, say X_3^T at a hypothetical near-horizontal material surface at a height sufficiently removed from the levels of interest to prevent effects of the erroneous estimation of the elevation of this surface from being serious. In the Lagrangian view, X_1, X_2, X_3 and t are the independent variables of the meteorological equations and the dependent variables include x_1, x_2, x_3 In addition we introduce the dynamical variables:

$$\pi = \left(p/p_{ref}\right)^{\kappa} \equiv \text{Exner pressure function}$$

where p is the pressure and p_{ref} is some constant reference pressure. Also

$\kappa = 2/7 = (\gamma-1)/\gamma \equiv$ Poisson's constant.

$\theta = T/\pi \equiv$ potential temperature

the density ρ is then

$$\rho = R(p_{ref})^{-1/\kappa}\ \pi^{\frac{1}{(\gamma-1)}}\ \theta^{-1} \qquad (15)$$

A material element of the fluid occupies a constant 'volume' of (X_1, X_2, X_3)-space and, assuming Lagrangian conservation of mass, the quantity, α defined as the mass per unit 'volume' of (X_1, X_2, X_3) – space is a static distribution for any given set of Lagrangian coordinates. We call this quantity the 'coordinate-density' and note that it is related to physical density by

$$\alpha = J\frac{\rho}{a^2} \qquad (16)$$

where J is the Jacobian,

$$J = \det\left[J_{ij}\right] \qquad (17)$$

and $$J_{ij} = \frac{\partial x_i}{\partial X_j} \qquad (18)$$

Adopting the shorthand,

$$\dot{c} \equiv \frac{dc}{dt} \quad ; \quad \ddot{c} \equiv \frac{d^2c}{dt^2}$$

for Lagrangian derivatives of arbitrary field, c, then the dynamical equations may be grouped as follows:-

$$\left.\begin{aligned} \ddot{x}_1 &= \alpha(F_1 + G_1) \\ \ddot{x}_2 &= \alpha(F_2 + G_2) \end{aligned}\right\} \qquad (19)$$

where G_1 , G_2 are defined in (14) and

$$\left.\begin{aligned} F_1 &= \frac{1}{a} f \dot{x}_2 - a c_p \theta \frac{\partial \pi}{\partial x_1} + \left\{ \text{Friction} \right\}_1 \\ F_2 &= -\frac{1}{a} f \dot{x}_1 - a c_p \theta \frac{\partial \pi}{\partial x_2} + \left\{ \text{Friction} \right\}_2 \end{aligned}\right\} \quad (20)$$

$$\frac{\partial \pi}{\partial x_3} = -\frac{g}{c_p \theta} \quad (21)$$

$$\dot{\theta} = \frac{\left\{ \text{Heat} \right\}}{c_p \pi} \quad (22)$$

$$\dot{\alpha} = 0$$

ie $$\alpha(\underset{\sim}{X}, t) = \alpha(\underset{\sim}{X}, t_0) \quad (23)$$

$$\alpha = \frac{J R (\rho_{ref})^{-1/\kappa} \pi^{\left(\frac{1}{\gamma-1}\right)} \cdot \theta^{-1}}{a^2} \quad (24)$$

with J defined in (17).

In a non-hydrostatic model the 'hydrostatic equation' (21) would be replaced by

$$\dot{x}_3 = -g - c_p \theta \frac{\partial \pi}{\partial x_3} \quad (25)$$

Finally, we wish to replace all partial derivatives $\frac{\partial}{\partial x_i}$ from equations (19) to (24) by derivatives with respect to the independent variables, ie terms of the form $\frac{\partial}{\partial X_i}$ This is done by application of the 'chain-rule':

$$\frac{\partial}{\partial x_i} = \sum_{j=1}^{3} J^{-1}_{ij} \frac{\partial}{\partial X_j} \qquad (25)$$

It is therefore possible to express all the spatial derivatives in the governing equations in terms of first-derivatives with respect to the Lagrangian coordinates, X_i . Discussion of the friction and heating terms is left until section 5 which is the discussion of processes involving moisture.

Possible boundary conditions are:

$$x_3(X_1, X_2, 0) = Z_0(x_1, x_2) . \qquad (27)$$

where Z_0 is the topographic elevation

$$x_3(X_1, X_2, X_3^T) = Z_T \qquad (28)$$

where Z_T is the height of the rigid lid. The upper condition, (28) could no doubt be refined to prevent spurious reflection of disturbances.

3. Numerical representation in space

Since the accuracy of the dynamics crucially depends on the accuracy of the spatial finite differencing, and in particular, on the coefficients J^{-1}_{ij} in (26) it is appropriate that we endeavour to maintain a Lagrangian computational grid on which the variation of the x_i is smooth with respect to the X_j . If this can be achieved, we may be confident that, by using high-order finite differencing formulae the transformation of the other differential quantities from "gradients in $\underset{\sim}{X}$" to "gradients in $\underset{\sim}{x}$" will be done with negligible loss of accuracy. To this end it is proposed to introduce periodically a selective numerical filter of the dependent variables (especially $\underset{\sim}{x}$) to remove structures that are of a scale in $\underset{\sim}{X}$ - space approaching the grid resolution (for example, less than 4 grid lengths). It is possible to devise multipoint filters that have this effect.

In a Lagrangian grid there is almost certainly nothing to be gained by a staggering of the different dynamical variables so a single computational grid

of points varying smoothly in $\underset{\sim}{X}$ is recommended. Within this restriction of smoothness it would probably be desirable to exploit the great versatility of the Lagrangian formalism of the dynamics in order to alleviate some of the other practical problems of implementation, such as the requirement to overlap efficiently the transfer of data between core and backing-store with the computations on the data in core. Since high-order spatial differencing requires broad regions of the dynamical fields to be in core simultaneously it is better to scan the data in an order that naturally tends itself to the holding, and continuous transferring, of well-proportioned 'blocks' as opposed to long narrow 'strips'. In this respect a helical grid arrangement, as illustrated in Figure 3, is better than a more conventional 'latitude-longitude' arrangement. Topologically it would consist of a single long ribbon of successive rows of data, each row (across the ribbon) containing of the order 10 to 30 grid points and the ribbon wrapped as many times around itself as is required to cover the sphere with the grid-resolution chosen, like the continuous peel of an apple. The two almost-antipodal grid-singularities would (as with all grid singularities) require special treatment. Away from these points the transfer of data to and from backing-store would be along 'fronts', A and C, spanning 3 or 4 widths of the ribbon as indicated in Figure 3 and would allow the application of high-order differences (involving broad spans of grid points in each direction) to update the points in the central part of region B that is temporarily in core. It should be noted that with this arrangement of data the resolution of a global model is limited not by the available core-store but rather by the total available backing-store that can be accessed while computations proceed. The latter quantity is generally many times the former and this will justify the adoption of numerical schemes that are extravagant on (temporary) storage in order to be economical on computation and very accurate in the final results.

4. Numerical integration on time

Having stated in the introduction that one of the most important potential benefits of a Lagrangian formulation is the ability to perform the time-integration with fewer (ie longer) steps we should discuss here the kind of numerical methods

that are suitable for this problem. In conventional forecasting models the timesteps (a few minutes) are so much shorter than the timescales of primary interest (a few hours) that little accuracy is lost by using methods that are formally of a low order of accuracy in time. This is no longer true for a Lagrangian model with a relatively long time step so in order to minimize time-truncation error it will be necessary to construct methods of integration that possess a high order of accuracy in time. Also, the existence of feedback timescales much shorter than the periods of interest requires that the methods of integration are equipped to handle these rapid feedbacks stably (but not accurately since rapid modes do not appear to be meteorologically significant on the synoptic scale). What are required are high-order-accurate implicit methods of integration, a sufficiently general class of which will now be described.

We start by grouing the dynamical variables into 'prognostic' variables (x_1, x_2, θ) and 'diagnostic' variables, (x_3, α, π). In principle, if the three dynamical modes per grid point (one meteorological mode, two gravity modes) are uniquely determined by the evolution of variables x_1, x_2, θ of the prognostic set, then the remaining variables, x_3, α, π may be determined from them by the diagnostic equations (21), (23) and (24). It is therefore necessary to use accurate time-integration methods for x_1, x_2, θ only since the other variables follow from them diagnostically. For algebraic convenience we shall identify the prognostic variables, (x_1, x_2, θ) by (ψ_1, ψ_2, ψ_3) and the diagnostic variables (x_3, α, π) by (ψ_4, ψ_5, ψ_6) respectively. We shall regard equations (19), (20), (22) as the 'prognostic equations' for (ψ_1, ψ_2, ψ_3) and equations (21), (23), (24) as the 'diagnostic equations'. Notice that the prognostic equations have terms explicitly containing time-derivatives of the prognostic variables. To implement the accurate implicit methods, first augment the prognostic variables with a number of their successive time-differentials,

eg

$$\psi_{i,j} \equiv \frac{d^j \psi_i}{dt^j} \qquad \begin{matrix} i = 1, 2, 3 \\ j = 0, 1 \ldots M \end{matrix} \qquad (29)$$

plus some auxillary variables,

$$\chi_{i,j} \qquad \begin{matrix} i = 1, 2, 3 \\ j = 1, \cdots N \end{matrix}$$

which record other characteristics of the evolution of the respective prognostic variables ψ_i . M determines the eventual order of accuracy of the scheme in time and N is empirically determined by the number of auxilliary variables required to maintain numerical stability.

In what we call the 'predictor step' we use this expanded set of variables to project the 'state' forward a time step, δt by means of a modified Taylor expansion:

$$\left.\begin{aligned} \psi^{0}_{i,j}(t+\delta t) &= \sum_{k=0}^{M} T_{jk}\,\psi^{L}_{i,k}(t) + \sum_{k=1}^{N} R_{jk}\,\chi^{L}_{i,k}(t) \quad \begin{pmatrix} i=1,2,3 \\ j=0,\cdots M \end{pmatrix} \\ \chi^{0}_{i,j}(t+\delta t) &= \sum_{k=1}^{N} S_{j,k}\,\chi^{L}_{i,k}(t) \quad \begin{pmatrix} i=1,2,3 \\ j=1,\cdots N \end{pmatrix} \end{aligned}\right\} \quad (30)$$

ie $$\left|\begin{matrix}\psi^{0} \\ \chi^{0}\end{matrix}\right|_{t+\delta t} = \left[\begin{array}{c|c} T & R \\ \hline 0 & S \end{array}\right] \left|\begin{matrix}\psi^{L} \\ \chi^{L}\end{matrix}\right|_{t}$$

where $(\psi^{L}, \chi^{L})_t$ are the prognostic state-variables at time t used for this 'predictor' step and $(\psi^{0}, \chi^{0})_{t+\delta t}$ are the resulting state variables thus predicted at time $t+\delta t$. The matrix T_{jk} is the constant Taylor-coefficient transfer matrix;

$$\left.\begin{aligned} T_{jk} &= \frac{(\delta t)^{(k-j)}}{(k-j)!} \qquad k \geq j \;,\; j \geq 0 \\ T_{jk} &= 0 \qquad\qquad k < j \end{aligned}\right\} \quad (31)$$

and S and R are constant matrices empirically optimised for the typical

feedback characteristics of the model.

The initial prediction $\psi^{0}_{i,j}$ is clearly obtained without any explicit dynamical information and in general will need to be corrected in order that the governing equations are satisfied. The remaining (diagnostic) variables at time, $t+\delta t$ may be iteratively estimated from the diagnostic equations (using their values for time t as a starting guess).

The 'predictor step' is followed by the 'corrector step'. The error-residuals are calculated from the prognostic equations applied at each grid point to $\underset{\sim}{\psi}^{0}$ (ie substituting $\psi_{i,1}$ for $\dot{\psi}_i$ and $\psi_{i,2}$ for $\ddot{\psi}_i$). Then a set of corrections are iteratively inferred from the magnitudes and spatial distribution of these residuals to refine the estimates of the variables at $t+\delta t$ in a way strictly constrained to be of the form:

$$\left.\begin{aligned} \psi^{n+1}_{i,j}(\underset{\sim}{X}, t+\delta t) &= \psi^{n}_{i,j}(\underset{\sim}{X}, t+\delta t) + q^{n}_{i}(\underset{\sim}{X}, t+\delta t)\, P_j \\ \chi^{n+1}_{i,j}(\underset{\sim}{X}, t+\delta t) &= \chi^{n}_{i,j}(\underset{\sim}{X}, t+\delta t) + q^{n}_{i}(\underset{\sim}{X}, t+\delta t)\, Q_j \end{aligned}\right\} \begin{matrix} 0 \le n \le L-1 \\ i = 1,2,3 \end{matrix} \qquad (32)$$

where (P_j, Q_k) are a pair of constant vectors, also empirically estimated for the characteristics of the model, and n is the index of this iteration. Associates with each iteration of these prognostic variables is a recalculation of the diagnostic variables. The correction components, $q_i(\underset{\sim}{X}, t+\delta t)$ at each grid-point will be derived from the residuals of the prognostic equations typically by solving elliptic equations similar in structure to those that appear in diagnostic balance constraints for filtered models. The particular techniques used to derive those corrections are not crucial to the nature of the integration technique so long as the successive state variables,

$$\psi^{n}_{i,j} \quad (i = 1, 2, 3) \quad \text{and} \quad \psi^{n}_{k} \quad (k = 4, 5, 6)$$

conform to the governing equations with progressively diminishing residual error as n increases. However, the rate of convergence of this iteration to determine

the total corrections, $\sum_n q_i^n(\underset{\sim}{X}, t+\delta t)$ might be greatly accelerated by adopting the very efficient modern relaxation-methods such as the 'multigrid' technique, Brandt et al (1980).

A vital feature of these methods is that it is not necessary to iterate the corrections to perfection in order to guarantee an accurate and stable overall prediction. It is merely sufficient to iterate the corrections enough to impose the information inherent in the governing equation upon the 'state' comprising the current dynamical variables and other computational variables. It is the predictor-steps rather than the corrector-steps that ensure a high formal order of accuracy and for this reason we should not underrate the importance of the extra computational variables since, in their relationship to one-another they implicitly carry valuable dynamical information.

The use of high order accurate implicit methods has a fairly long and well established history in the numerical solution of ordinary differential equations and a comprehensive introduction to the techniques is given in Gear (1971). Their application in the solution of partial differential equations is less well established although there is no impediment to their adaptation for use with the more complex systems of equations such as the governing equations used here. Their behaviour in more complex systems can be approximately analysed by separating out the eigenmodes of the numerical system using a relatively simple linearisation, thereby reducing the analysis to the form used for analysing ordinary differential equation methods.

The implicit method is uniquely specified by the order of accuracy, M the number of auxilliary variables, N per prognostic variable, the $N \times N$ matrix S (which, for stability, must have eigenvalues of absolute magnitude less than one), the $(M+1) \times N$ matrix R, the $(M+1)$ - element vector P and the N - element vector Q which determine the allowed ratio of corrections to the computational variables $\psi_{i,j}$ and $\chi_{i,\kappa}$ associated with prognostic variable ψ_i. By a canonical transformation of variables $\psi_{i,j}$ and $\chi_{i,\kappa}$ it is possible to reduce matrices S and R to relatively simple standard forms, greatly simplifying their theoretical analysis and optimisation. Beyond

that, there appears to be no elegant method of choosing the best set of values for these quantities and one must presumably resort to some form of computer-aided optimisation, controlled by the expected range of the dynamical system's feedback times and its timescales of meteorological interest, by maximising an appropriate composite measure of the scheme's accuracy (eg at different forcing frequencies and system-feedback times) subject to the overriding requirement of numerical stability. Dahlquist (1963) has proved that no high-order implicit methods of the form we have considered are stable for all possible linear stable feedback characteristics of the system. Nevertheless, for the finite range of feedback times (ie up to the most rapid gravity oscillation resolved spatially) in a finite model there do exist practical methods of the type described in general terms here that are adequate.

5. Unscrambling the grid (rezoning)

After a while any initially regular Lagrangian grid becomes hopelessly tangled and it is essential to renew the grid well before this disaster occurs. There are clearly many ways of interpolating data from one (slightly scrambled) grid to another (less scrambled) grid but to do this in a way that faithfully retains the information on the original grid requires some care. It would be almost pointless to mar an otherwise accurate integration of a Lagrangian model by a periodic interpolation that uses a crude and inaccurate algorithm (eg linear, or low order splines). Also it would be wasteful not to exploit the existing structural regularity of the old grid when interpolating to the new. We shall suppose that we wish to interpolate the data from the old $(\hat{X}_1, \hat{X}_2, \hat{X}_3)$ grid to the new $(\hat{X}_1', \hat{X}_2', \hat{X}_3')$ grid in a way that guarantees a high-fidelity reproduction of the physical information ($\hat{X}_i$ denotes the coordinates of the grid surfaces).

While it is quite feasible to interpolate to each $(\hat{X}_1', \hat{X}_2', \hat{X}_3')$ point directly from a set of $(\hat{X}_1, \hat{X}_2, \hat{X}_3)$ points using a polynomial method of formally high-order of accuracy, this method can become disproportionately expensive when the order of accuracy required becomes large. eg for a formula of

9th order accuracy (in one-dimension this means interpolating using a polynomial fitting 10 grid points) in each of the three directions the number of computations by the direct method would be at least 10 x 10 x 10 individual multiplications and additions for a single target-point. ie for a grid of N_g points that means about 1000 N_g multiplications and additions even in the artificial situation in which the interpolation coefficients are assumed to be already known. Clearly a method that can reduce this excessive computation would be advantageous. Suppose instead that the interpolation between the grids is conducted via intermediate grids (which we assume also have about N_g points) according to the three step scheme:

$$(\hat{X}_1, \hat{X}_2, \hat{X}_3) \longrightarrow (\hat{X}_1', \hat{X}_2, \hat{X}_3)$$

$$(\hat{X}_1', \hat{X}_2, \hat{X}_3) \longrightarrow (\hat{X}_1', \hat{X}_2', \hat{X}_3) \qquad (33)$$

$$(\hat{X}_1', \hat{X}_2', \hat{X}_3) \longrightarrow (\hat{X}_1' \hat{X}_2' \hat{X}_3')$$

Each step now involves about $10 N_g$ multiplications and additions, ie only about $30 N_g$ for the entire sequence. Thus by interpolating to the new grid via two 'hybrid' grids it is possible to achieve an accurate transfer of data quite efficiently. The difficult problem is determining the relative positions of the grids. This problem is made easier if, as we shall assume, the 'standard' unscrambled $\underset{\sim}{X}'$ coordinates are analytic, or least readily evaluated, functions of position. Then, since all the $\underset{\sim}{x}(\hat{X}_1, \hat{X}_2, \hat{X}_3)$ are known (being dynamical variables) we may assume that

$$X_1'(\hat{X}_1, \hat{X}_2, \hat{X}_3)$$
$$X_2'(\hat{X}_1, \hat{X}_2, \hat{X}_3)$$
$$X_3'(\hat{X}_1, \hat{X}_2, \hat{X}_3)$$

are also known.

To interpolate from grid $(\hat{X}_1, \hat{X}_2, \hat{X}_3)$ to $(\hat{X}_1', \hat{X}_2, \hat{X}_3)$ it is required to know each $X_1(\hat{X}', \hat{X}_2, \hat{X}_3)$. From the list of values $X_1'(\hat{X}_1, \hat{X}_2, \hat{X}_3)$ on the appropriate line of constant $\hat{X}_2$ and $\hat{X}_3$ it is possible to make a sensible first estimate

$$X_1^{(0)} \simeq X_1(\hat{X}', \hat{X}_2, \hat{X}_3) \tag{34}$$

for a particular intermediate grid point. This estimate is imposed by the following scheme (Newton's method):

(i) For estimate $X_1^{(n)}$ derive the actual value of X_1' by the accurate interpolation

$$X_1'^{(n)} = X_1'(X_1^{(n)}, \hat{X}_2, \hat{X}_3) \tag{35}$$

(ii) Compare this with the intended value, $\hat{X}_1'$, and infer a correction to the current estimate, X_1 :

$$X_1^{(n+1)} = X_1^{(n)} - \frac{X_1'^{(n)} - \hat{X}_1'}{\left(\frac{\partial X_1'}{\partial X_1}\right)_{approx}} \tag{36}$$

where $\left(\frac{\partial X_1'}{\partial X_1}\right)_{approx}$ is a simple differenee approximation to $\frac{\partial X_1'}{\partial X}$ (along constant $\hat{X}_2$ and $\hat{X}_3$) in the appropriate interval of the X_1 - grid.

(iii) Repeat (i) and (ii) as many times as is required to obtain convergence. When $X_1(\hat{X}_1', \hat{X}_2, \hat{X}_3)$ is known for all such points it is possible to carry out the first step of (33) for any variable available on the old $(\hat{X}_1, \hat{X}_2, \hat{X}_3)$ grid. We interpolate X_2' and X_3' so that

$$X_2'(\hat{X}_1', \hat{X}_2, \hat{X}_3)$$

and $$X_3'(\hat{X}_1', \hat{X}_2, \hat{X}_3)$$

are known. The procedure (34), (i), (ii), (iii) is then applied to the next step, to obtain,

$$X_2(\hat{X}_1', \hat{X}_2', \hat{X}_3)$$

and hence (by interpolation):

$$X_3'(\hat{X}_1', \hat{X}_2', \hat{X}_3)$$

Finally, by the same procedure, we iteratively get

$$X_3(\hat{X}_1', \hat{X}_2', \hat{X}_3')$$

so that the sequence of interpolations (33) may now be executed for all the dynamical variables.

The algorithm outlined above will only work if the various grids intesect each other in a well behaved way, so that the intersections required for process (33) are unique and distinct. In the viscinity of the grid singularities we will be forced to use special methods, probably involving the interpolation first to some locally regular grid.

6. Treatment of sub-grid-scale processes and embedded structures

In this section we examine the special difficulties and, in some cases, opportunities provided by a Lagrangian formulation for the treatment of physical and dynamical processes on scales too small to be accurately resolved by the computational grid. The obvious difficulty with the treatment of manifestly anisotropic processes such as convection, boundary-layer friction, radiative transfer and vertical turbulent heat transport is the loss of the vertical alignment of grid points that is taken for granted in an Eulerian model. This means that physical parametrisations must be formulated in a particularly careful way in order that they be applicable in a model which at any given time has an effectively general curvilinear grid-coordinate system.

The sort of approach that seems most likely to succeed in a Lagrangian frame is one that expresses the relevant physical processes in a way that is formally

invariant with respect to coordinate change. Vector (and tensor) calculus epitomises this property and suggests that if the forcing of the resolvable dynamics by physical processes can be expressed entirely by the divergences of fluxes (vector or tensor fields) then the problems of physical parametrisation may be satisfactorily solved. The fluxes themselves should in turn be coordinate-invariant diagnostics of the local flow. Certainly processes of a diffusive nature (friction, turbulent heat transport in the boundary-layer, etc) can indeed be characterised in flux-form and, provided the fluxes themselves are temporarily held as explicit variables of the model, the equations involve no more than first derivatives in space.

eg. making Reynold's stress components, τ_i physical variables (held at Lagrangian grid points), we have

$$\{\text{Friction}\}_i \propto \frac{\partial \tau_i}{\partial x_3} \qquad i = 1, 2 \qquad (37)$$

with $$\tau_i = K \frac{\partial(\dot{x}_i)}{\partial x_3} \qquad (38)$$

($K \equiv$ diffusivity)

which, by application of the 'chain-rule' can be simply expressed in terms of the Lagrangian coordinates.

The representation of convective activity by a flux-field might also be feasible. The effects that are significant dynamically are those that result in a transport of heat, mass, moisture and possibly momentum. By incorporating additional physical variables at each grid point to represent the magnitudes of these (upward) convective fluxes and modifying the mass continuity equation (23) to allow for the divergence of 'convective mass-flux' (ie a source term at the level of the top of convective updraughts and a sink at the bottom) it seems quite plausible that the statistical results of many convective updraughts can be properly accounted for. With additional fluxes, downdraught effects may be dealt with in a similar way. These fluxes would then obey their own diagnostic

equations, presumably involving each other, the vertical gradients of temperature and momentum, and humidity.

Radiative processes can clearly be written in flux-form, the additional physical variables of the scheme being in this case the upward and downward radiant flux intensities of an appropriate number of spectral bands. Again, the vertical divergences of these fluxes would force the large scale dynamical variables.

While this tentatively proposed approach to physical parametrisation requires several additional 'physical' variables, they would usually be diagnostic variables and therefore would not necessarily need to be retained when the fields are written to backing-store. However, to avoid the possibility of numerical instability and to be consistent with the dynamics these parametrisations should be made 'implicit' together with the dynamics, and hence iterated each time step.

Where the intrinsic characteristics of a Lagrangian scheme can be most profitably exploited is in the automatic advection of explicit detailed sub-grid-scale structures whether they be 'physical' features such as rain bands, strato-cumulus sheets, cloud clusters etc. or 'dynamical' structures such as fronts, squall lines, mesoscale vortices depicted in a passive or semi-passive way by an advected pattern of subgrid-scale potential vorticity. One possibility is to construct a 'map' of embedded structures on a relatively fine grid formed by subdividing the Lagrangian computational grid, and using informative quasi-conservative variables such as potential vorticity, total humidity mixing ratio etc and to allow this map to be transported automatically by the large scale dynamics with the embedded variables updated only infrequently and with only the crudest (ie computationally cheapest) additional equations. An obvious 'source' for this fine scale structure is the grid-scale component of flow deliberately filtered from the main Lagrangian grid (see section 3) in order to maintain its smoothness. The fine scale variables would of course be subject to the rezoning operations but not by such accurate methods as those suggested in section 5 since these fine scale patterns are not sufficiently reliably known (even initially) to justify that expense.

The embedded structures will be especially useful for short range and mesoscale forecasts although in the absence of proper 'dynamics' for these features they might require rather careful interpretation. The ability to advect relatively fine-scale potential-vorticity structure at minimal computational expense is potentially of immense value since at appropriate forecast times (eg every six hours) it would be possible to iteratively invert simplified (elliptic) equations of balance forced by this field in order to infer the corresponding detailed fields of wind, pressure and vertical velocity (without requiring these quantities to be held and continuously updated through intervening times). Also, with a little more effort, the detailed winds thus obtained could be applied to update (by advection) the detailed potential vorticity pattern to allow for the self-interaction of the small scale flow. With such a 'two-tier' configuration of dynamics based on a coarse-scale Lagrangian model carrying embedded structure it might well be possible to achieve high-resolution forecasts at a computational cost and at a standard of accuracy that could not be matched by conventional methods.

7. Conclusion

A radical approach to the problem of numerical weather prediction is suggested, based on the treatment of dynamics and physics in a Lagrangian framework and numerical techniques of space-differencing, time-integration and grid-rezoning that can be readily extended to arbitrary order of accuracy so as to minimize all forms of numerical truncation and interpolation errors. An essential component of the proposed approach is that the governing equations are solved in a fully implicit way, thus avoiding the possibility of linear numerical instability which, in various forms, still continues to plague conventional models. The growth of structures of a scale barely resolved by the computational grid would be periodically controlled by application of selective low-pass numerical filters.

The proposed approach would require extensive backing-store (discs) because of the extra computational variables required to achieve very accurate and stable integration in time with relatively long time steps. But it should be relatively economical on computation considering the resolution and standards of accuracy that can be expected for the technique. While it might now be premature to attempt

to construct a full multi-level Lagrangian model along the lines presented here, it would be of value to investigate the performance of the proposed methods on a simpler system such as the inviscid shallow-water equations in a small rectangular domain with rigid walls or cyclic boundary conditions.

References

Brandt, A., Dendy, J.E., Ruppel, H. (1980) The multigrid method for semi-implicit hydrodynamic codes. J. Comp. Phys. 34 pp. 348-370.

Christiansen, (1970) VORTEX, a 2D hydrodynamic simulation code. Culham Lab. Report. CLM-R-106.

" (1973) Numerical simulation of hydrodynamics by the method of point vortices. J. Comp. Phys. 13 pp. 363-378.

Crowley, W. P. (1968) Numerical advection experiments. Mon. Wea. Rev. 96 pp. 1-11.

Dahlquist, G. G. (1963) A special stability problem for linear multistep methods. BIT 3 pp. 27-43.

Gadd, A. J. (1978) A numerical advection scheme with small phase errors. Quart. J.R. Met. Soc. 104 pp. 583-594.

Gear, C. W. (1971) Numerical initial value problems in ordinary differential equations. Prentice Hall, New Jersey.

Mesinger, F. (1971) Numerical integration of the primitive equations with a floating set of computational points: experiments with a barotropic global model. Mon Wea. Rev. 99 pp. 15-29.

Robert, A. (1981) A stable numerical integration scheme for the primitive meteorological equations. Atmos. Ocean. 19 pp. 35-46.

Robert, A. (1982) A semi-Lagrangian and semi-implicit numerical integration scheme for the primitive

meteorological equations.
J Met Soc. Japan. 60 pp 319-325.

Sawyer, J. S. (1963) A semi-Lagrangian method of solving the vorticity advection equation. Tellus 15 pp. 336-342.

Wiin-Nielsen, A. (1959) On the application of trajectory methods in numerical forecasting. Tellus 11 pp. 180-196.

Figure Captions

Figure 1. Advection of an impulsive 'spike' across 20 grid lengths by a uniform flow with CFL number, $U \delta t / \delta x = 1/4$ using the following methods: (a) exact; (b) upstream differences; (c) centred leapfrog; (d) Gadd's scheme.

Figure 2. Diagram to show the effect of the polar stereographic projection.

Figure 3. Part of a helical grid system showing zones at which data are being read out (A), used for computation (B), and read in (C).

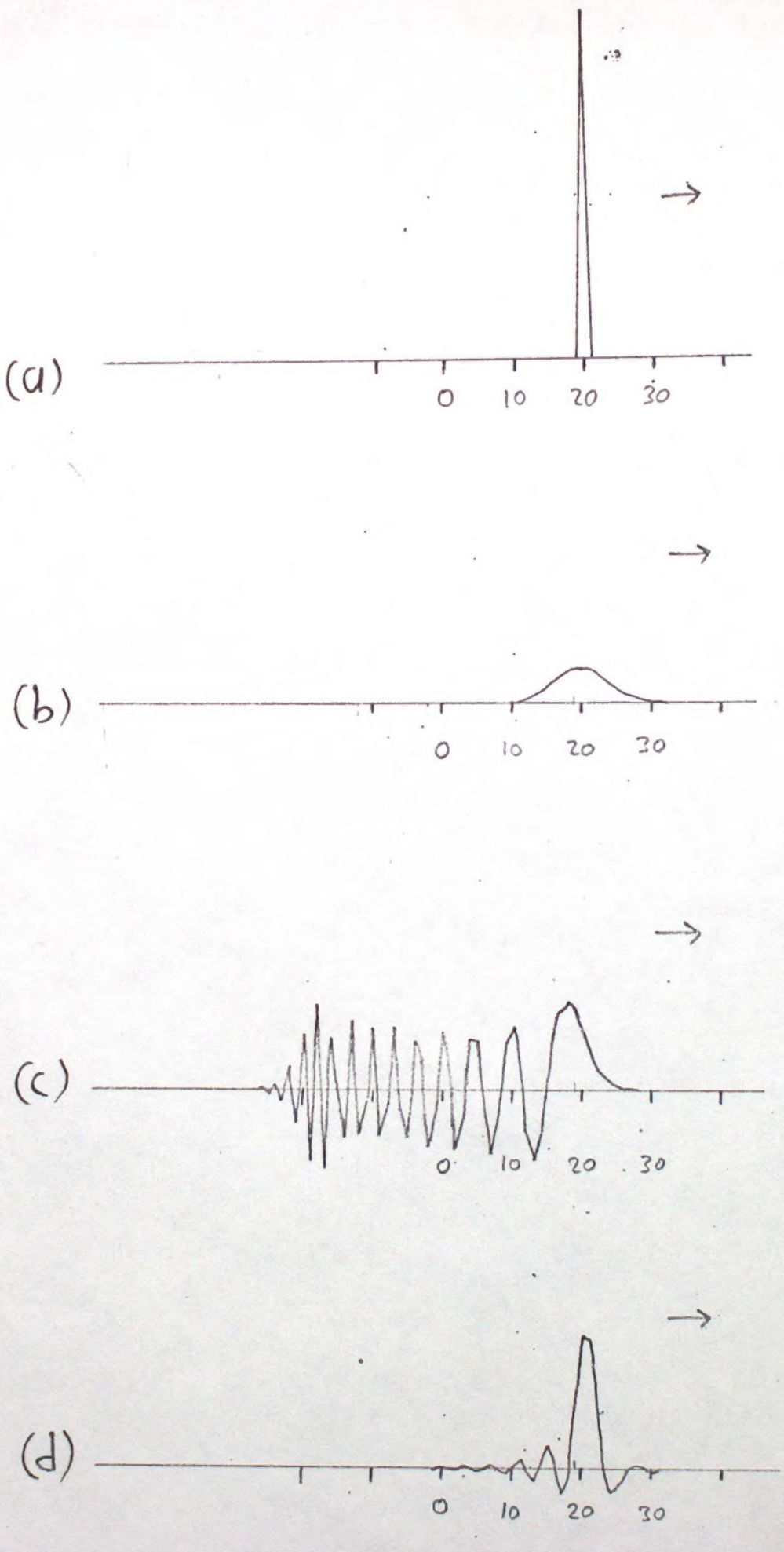

FIGURE 1

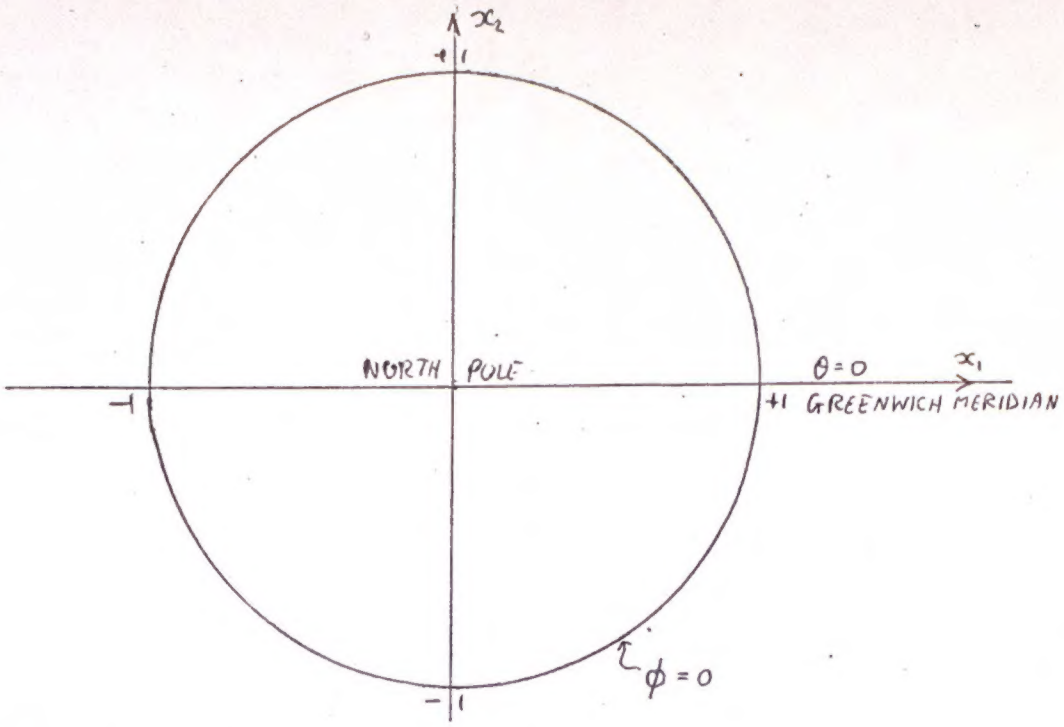

FIGURE 2

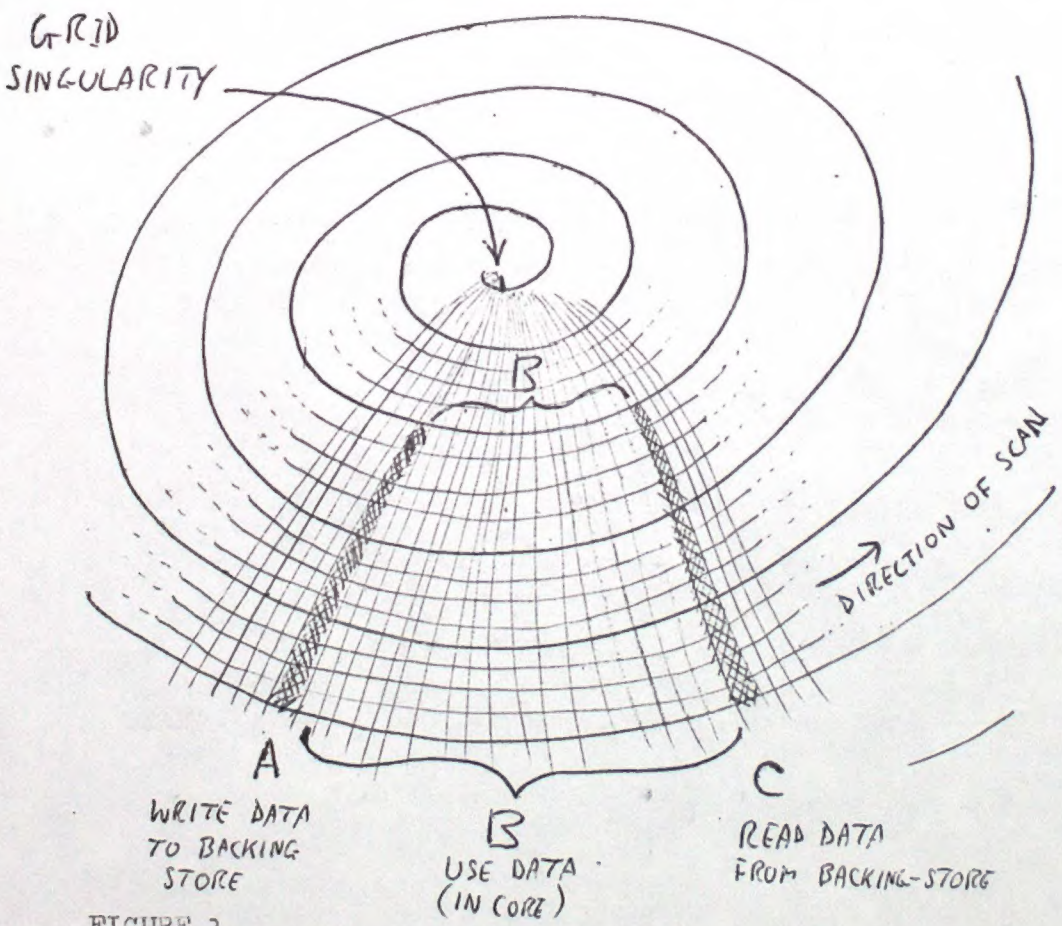

FIGURE 3